# SPATIAL DATA SYSTEMS: ORGANIZATION OF SPATIAL DATA

by

Kenneth J. Dueker

TECHNICAL REPORT NO. 4

of

ONR Task No. 389-143
Contract Nonr 1228(37)

OFFICE OF NAVAL RESEARCH

GEOGRAPHY BRANCH

Department of Geography
Northwestern University
Evanston, Illinois

December 1966

Funding Provided by
Natural Resources Program
Office of Space Sciences and Applications
National Aeronautics and Space Administration

# PREFACE

This report provides a limited explication of current needs for classifying and organizing spatial data for use in urban and transportation planning. In addition, requirements and methods for handling spatial data are explored. Mr. Dueker emphasizes the dual need for data organization methods and data handling capabilities, as a requisite for utilization of data acquired from remote sensors mounted on earth orbital platforms. This work provides a basis for examination of some problems of integrating remote sensors into a viable geographic information system.

SPATIAL DATA SYSTEMS

by

Kenneth J. Dueker

ABSTRACT

Spatial data systems are concerned with the organization, handling, and retrieval of data whose spatial position is of concern. Spatial data are of particular concern in urban and transportation planning. In these fields considerable attention is given to spatial and temporal variations of data.

The three volume report presents a discussion of concepts and techniques that are essential in moving towards flexible and responsive urban information systems. The following areas are emphasized:

1. Explication of terms associated with spatial data.
2. Discussion of means of organizing spatial data for flexible and efficient retrieval.
3. Investigation of data handling capabilities for organizing and manipulating spatial data.
4. Presentation of these topics in a tutorial form, conceivably to serve as a text where none presently exist.

The greater speeds and storage capacities of newer computers requires new concepts of data organization and new means to create and access these more complex data structures. Of particular concern in urban and transportation planning are needs to link separately collected data that relate to the same phenomena or spatial locations, and a need for user-oriented data

handling capabilities. These needs are explored and recommendations are made.

Volume 1 contains the Summary and Conclusions, Introduction, Data Handling Capabilities for Supplying the Demands for Urban Data, Nature of Spatial-Temporal Data, and Questions and Queries.

Volume 2 deals with a Language to Query Spatial-Temporal Data, a Directory of Spatial-Temporal Data, and Hardware-Software Configurations.

Volume 3 deals with Entity Linkage, Organization of Data for Retrieval, Search Efficiencies, Point in Polygon Procedures and Retrieval from Digitized Imagery.

# Part I

## SUMMARY AND CONCLUSIONS

Spatial data systems are concerned with the organization, handling, and retrieval of data whose spatial position is of concern. Spatial data are of particular concern in urban and transportation planning. In these fields considerable attention is given to spatial and temporal variations of phenomena, such as land use and population characteristics. Typically, these variations are observed as differences in areal aggregates, such as differences in population densities or auto ownership rates for regions or areal units. More sophisticated mathematical modeling of urban structure is requiring more detailed spatial data; more detailed in locatability and more detailed frequency of observation. These data demands are putting great pressure upon providing or supplying data. New methods of organizing and handling data are essential.

### Characteristics of Spatial Data

For use in urban and transportation planning it is imperative that spatial position of data be observed. A grid coordinate system, tied to a known latitude and longitude is recommended for defining spatial position. A coordinate system provides a ratio scale of measurement, and distances may be calculated between observations. Conversely, if codes for areal units are used for defining spatial position, a map must be used to interpret spatial relationships. This is because locational codes are measurements on a nominal scale.

There is a need for agreeing upon a terminology concerning data and a need for classifying data. For the purpose of this investigation such a terminology relating to data is developed and data are classified by spatial-temporal properties. Here, data are defined as collections of values of selected properties for entities. For instance, the weight of a person is an item of data. The weight measure is the value, weight is property of a person, and the person is the observed entity. Spatial data relate to those phenomena observed in space and time and are classified as content data and place data. Content data have elements of phenomena, location, and time. Content data may be classified as observations on objects, events, activities and flows. Place data relate to the geometric characteristics of segments of space. These definitions aid in developing general schemes for organizing spatial data.

Data Handling

Existing data handling capabilities do not give enough consideration to spatial and temporal properties of data. Existing data systems treat spatial and temporal properties in the same way as other variables. Because data needs or questions in urban and transportation planning usually express or imply limits of location and time, it is imperative that spatial and temporal properties be used to organize data into files or sets. This means that a universe of like data observations, say observations on persons, are grouped into sets according to similar spatial-temporal properties. Observations possessing like location values and time values are grouped together into sub-files.

A generalized capability for handling spatial data consists of means to pose queries and to manipulate data. Queries are defined as search criteria in a form that are relatable to available data. To bring data to bear upon answering questions, the questions must be translated to queries. Although work on automated question-answering systems appears promising, the state-of-the-art is confined to very specific sets of data and specific classes of questions. Hence, the translation process is largely a human function. This study focuses upon the development of capabilities to query, rather than to question, diverse kinds of data. Thus a language to Query Spatial-Temporal (Quest) data is specified for subsequent development. Quest is designed to facilitate queries to and manipulation of spatial-temporal data, with built-in features for testing all entities within a file without special looping instructions and a built-in point-in-polygon test for determining whether entities are within query area(s). Implementation of Quest requires writing a translator program to translate Quest statements to some other operational programming language or to machine language.

## Directory System

To access data relating to specific locations and times, data should be organized into files according to their spatial-temporal properties. Accessing these files requires a directory mechanism or a known relationship between spatial-temporal values and the appropriate data set. Given as a key, a data type, a locational characteristic and a time value, the problem is to determine the type of data and the set possessing these spatial-temporal characteristics. This requires a data directory or index, or a mapping function that calculates the location of data given the spatial-temporal characteristics.

## Computer Systems

Clearly, hardware considerations are not independent of software, nor is software decisions independent of hardware.* The design of a spatial data system is the determination of a hardware-software configuration that performs the desired functions.** There are numerous possible hardware-software configurations. The problem is to define one's needs and objectives so that configurations can be devised and analyzed for selection of the "best" one. This study merely discusses hardware-software relationships along a continuum of increasingly sophisticated hardware.

## Organization of Data

To avoid reliance upon ecological correlation in urban and transportation planning analyses, it is necessary to organize data to enable the development of direct relationships between phenomena. Ecological correlation is often used to infer relationships between sets of aggregated areal data.

For example, when looking at individual households, there is a strong relationship between number of trips per household and household size. In trip generation analyses, averages or areal aggregates are often used, and the traffic analysis zones are assumed homogeneous. These analyses focus on variation between traffic analysis zones with the homogeneity assumption setting within-area variation equal to zero. Because household size does not vary extensively between areas, it is often found to be insignificant for predicting the number of trips per household. This and similar problems occur when using ecological or between-area correlation, and ignoring the within-area correlation.***

---

*Hardware is a term designating the mechanical, magnetic, electrical and electronic devices. Software are the computer programs as contrasted with computer components or hardware.

**A hardware-software configuration comprises the computer system element of a data system. A data system is a man-machine system consisting of a computer system, a data base, and systems technicians to interface with users.

***See Duncan, Cuzzort, and Duncan [1, p. 62] for a discussion of aggregation of areal data. (The number in brackets refers to entries in the list not references included at the end of each chapter.)

To avoid reliance upon ecological correlation and yet enable analysis of variations between areas or over space is a goal in urban transportation planning research. Achieving this goal requires greater care and skill in organizing data. For example it may require that separately collected data, relating to the same phenomena or location, be linked. This linkage of entities enables direct relationships to be investigated rather than relying upon inferences drawn from unrelated data.

The problem is to organize spatial data in forms that are amenable to linkage, stratification, and aggregation. All too often, the available forms of data restrict or control the analysis. Data are immediately aggregated to minimize the amounts of data and thus limit all subsequent data analysis. This concern with means of organizing and handling spatial data is significant in that the newer computers offer considerably more speed and direct access storage. These features provide greater opportunities for more extensive and detailed analyses. To take advantage of these greater speeds and greater storage capacities, new concepts of data organization are needed. In addition, systems like Quest are needed to create and access these more complex data structures.

In this context, Part IV deals with means for organizing and accessing spatial data. Linkage of separately collected data that relate to the same phenomena or spatial locations is similar to relating complex hierarchical data sets, such as household travel behavior data. Linking a trip to a person and the person to a household in what are called list structures is necessary for accurate representation of data and to eliminate initial organization of data to limiting preconceived forms that are not responsive to subsequent needs. One of the more important of the findings of this investigation is that data organization will become more important, for the storage of data in complex but flexible forms enables response to a wide variety of problems and analyses. However, to make these data organization concepts attractive and feasible, there must be a data handling capability, such as Quest, to create and update these complex data structures, as well as for retrieval of data from them. Successful data handling capabilities will be those that can do both. The data handling capability and data organization are interrelated, with the important aspect being flexibility. A flexible form of data is necessary as is a flexible data handling capability.

In conclusion, it appears safe to say that spatial data systems for urban and transportation planning are at an embryonic stage of development. "Information System" efforts and "Comprehensive urban models" fall short of expectations, primarily because of the lack of flexibility in organizing and handling spatial data. It is essential that future systems possess a greater capability for accessing and manipulating spatial data. This requires greater care and planning in the organization of data.

## Contributions

The contributions of this research effort are difficult to assess. Primarily this is because the investigation is somewhat unorthodox, with various conceptual materials being presented in a tutorial form. The necessity for this approach stems from a lack of awareness among professionals in urban and transportation planning of the potentials and limitations of computer technology applied to spatial data. The lack of awareness that spatial data differs from, say business data, is itself a major problem. Essentially, there are four major areas of contribution:

1. Explication of terms associated with spatial data.

2. Discussion of means of organizing spatial data for flexible and efficient retrieval.

3. Investigation of data handling capabilities for organizing, manipulating, and updating spatial data.

4. Presentation of these topics in a tutorial form, conceivably to serve as a text where none presently exist.

The identification and emphasis of these areas of concern is of more lasting significance than their treatment. This emphasis may focus more effort in these directions. With the development of cheaper and larger direct access storage devices and more sophisticated programming languages, and on-line computers, it is possible to store and have access to large amounts of data. To fully utilize this technology it is imperative that these four areas be fully exploited. The treatment given here is of an introductory nature. For example, Quest or something similar must be implemented and there must be numerous experiments in organizing spatial data.

The four contributions are considered to be equally important at this time. However, the areas described in items 2 and 3 will require the greatest amount of future effort. To achieve truly flexible spatial data systems considerable effort must be given to data organization and data handling capabilities. These two areas are essential. The data organization concepts discussed here and the design of Quest are illustrative of the needs in these areas.

## Limitations

The major limitation of this investigation is the lack of evidence to substantiate the assertion of the desirability of storing data in list structures. No empirical test was performed to test the supposed efficiencies of storing data in list form. A valid question is whether the extensive organization of data is warranted in comparison to a "brute force" alternative of testing all possible data.

Another limiting factor in weighing the contribution of this investigation is that a trial Quest translator was not written. Without actual experience in running Quest programs, it is difficult to assess its value.

So little has been done in studying problems of data organization and data handling with regard to urban and transportation planning. This investigation is therefore broad and general. Subsequent more detailed work must deal with the needs and efficiencies of data organization and data handling.

## References

1. Duncan, Otis Dudley, Ray P. Cuzzort, and Beverly Duncan. Statistical Geography, Free Press, 1961.

## Part II

## ORGANIZATION OF SPATIAL DATA

### CHAPTER 1. INTRODUCTION

#### Objectives and Scope

This study is specifically concerned with the storage and retrieval of data on spatially distributed phenomena. The storage and retrieval of spatially distributed data implies a capability of accessing recorded values of phenomena having locational characteristics. The emphasis is upon the selecting or accessing of stored data values rather than developing an informational meaning from context. The study focuses upon developing concepts for storing and retrieving data on spatially distributed phenomena. In particular, the study is concerned with storage and retrieval of data that are useful in the analysis of urban areas for planning purposes or a spatial data system for urban and transportation planning.

The objective is to conceptualize the elements of spatial data systems rather than to design one. The elements of a spatial data system consist of data inputs, representation of spatial data in storage, a data handling capability, and an output system for presenting spatial data. This study is primarily concerned with representation of spatial data and the conceptual development at generalized data handling capabilities.

#### Audience

The intent is to provide an interface between persons interested in the applications area, i.e., the analysis of urban areas, and persons interested in computer techniques. In provision of this bridge between applications and techniques, an attempt is made to provide enough detail in both areas for persons from either area. However, the study is probably of greater interest to, and is oriented towards urbanists who have some knowledge of and interest in computer technology.

## Approach to Spatial Data Systems

The approach to the design of spatial data systems is to create a capability for accessing data on spatially distributed phenomena in unanticipated ways. For a system to have long-term and broad utility it must have sufficient flexibility to enable usage, i.e., retrieval or manipulation of data, in ways which it was not explicitly designed to handle. This means that user capabilities must exist to enable modification and manipulation of these data. In turn, this means data must be expressed and represented in general terms.

Normally, one approaches the design of an automated data system in one of two ways. The first is to inventory the kinds of data that are presently being collected and to incorporate these into a unified automated system with multiple functions. These functions essentially reproduce the reporting or processing functions of the prior systems. This approach is usually applied to operational needs of agencies such as automating the billing operation for a utility service.

The second way to approach the design of an automated system is to enumerate the questions that are asked and to design a system to provide the answers. The key to this approach is in the formulation of questions. These questions must be formally expressed in terms of queries or imperative statements and applied to appropriate data sets. A query, as a formally represented question, acts upon the data base eliciting a response. This response is a selection of entities or items. This illustrates an important aspect of the enumeration of questions approach. Only those questions that are relatable to data, or that can be expressed as queries to a data set, can be explicitly answered. Questions that are not directly relatable to data require subjective human processing.* The problem is to formulate questions or elements of questions in terms of queries to obtainable data.

A third way to approach the design of a spatial data system is to think neither in terms of explicit data nor of explicit questions, but to conceive of generalized capabilities. These generalized capabilities relate to methods for handling classes of data and to answering classes of questions. This third approach is the one used in this study.

Organization of the Study

This study is divided into four parts. Part I is the summary and conclusions. The second part consists of this introduction, a review of existing data handling capabilities, the development of a framework for spatial-temporal data, and a differentiation between questions and queries. This second part is written for a fairly broad cliental with a minimal amount of technical detail being included. The primary purpose of Part II is to view data in the abstract terms and to demonstrate the need for generalized data handling capabilities.

The third part, with the exception of some terminology, is independent of the first for a person with knowledge of computer techniques. Part III is considerably more technical than Part II. Part III consists of a preliminary design for a query language, a formulation of a directory system to files of spatial-temporal data, and a discussion of considerations for the design and implementation of spatial data systems.

The fourth part consists of a variety of special topics that are of tangential interest to the main thrust of this study. These topics provide amplification for topics that are introduced in either Part II or Part III, and as such, are internally unrelated.

---

*In these situations a man-machine interaction may be necessary. For instance, a pseudo question may be asked of available data. This response is input to a subjective human process to answer the initial question.

Organization of Part II

Following this introduction, Chapter 2 presents a discussion of data supply and demand relationships and a review of some existing data processing systems. These systems are investigated for features that are applicable to the development of capabilities to enable a greater supply of spatial or urban data.

In Chapter 3 a general framework is developed for classifying data on spatially distributed phenomena. The importance of this part of the study is the classification of data by their spatial and temporal characteristics. These general classes of data provide a foundation to enable the design of systems to handle these classes of data.

Chapter 4 differentiates between questions and queries, briefly described in the previous section. The discussion of questions and queries provides the central focus for this study. Preceding topics are drawn together in Chapter 4, and it also serves as a point of departure for subsequent topics. The problem of query formulation necessitates the concern with the nature and organization of spatial-temporal data and with data handling capabilities.

Organization of Part III

In Chapter 5, a proposal for the Quest Language (QUEry of Spatial-Temporal Data) is presented. This presentation gives a defining description of the Quest language. Quest is suitable for expressing a large number of retrieval problems in a form sufficiently concise for direct automatic translation into the language of computers. Quest is a hardware independent language and is not designed for a particular computer configuration. Quest provides the opportunity for an immediate and short-run capability for handling data on spatially distributed phenomena. It is not proposed as the final answer, but as pressing needs in the near future.

Quest consists of five types of statements: a file selection or input statement, an entity selection statement, a modification statement, a manipulation statement and an output statement. With the judicious use of these statements, quite sophisticated programs can be written to act upon data. The input statement specifies the file or list to be searched. The entity selection statement essentially consists of an IF statement that enables the specification of conditions or values for selection. A modification statement acts as a replacement statement for the creation of new properties. The same replacement is made for each of the selected entities. A manipulation statement enables one to change the relationship of the entities themselves. Manipulation consists of summarizing or sorting the entities of the input data file. Finally, the output statement enables the reporting of the query response or graphic display in terms of spatially positioning data.

There are three important features to Quest. One is that the user-programmer writes his Quest program as if he were to read a single entity. That is, he does not have to be concerned with the details of instructing the computer to look at each entity in the file. The Quest translator generates the looping instructions for processing each entity of the program in the file. The second feature is that the Quest translator has a built-in point-in-polygon test which enables the determination of whether a spatially positioned entity is within an areal unit. The third feature is for graphic display of output data.

In Chapter 6 a directory system is considered that enables accessing data sets. Such a system is essential for a real-time capability on an on-line system. The purpose of a data directory system is to enable the efficient selection of files for querying. The diverseness of urban data, the spatial extent and the temporal extent pose problems in selecting data. Data are organized into files according to these spatial and temporal characteristics. Data are partitioned into files having uniform spatial regions and annual time periods, uniform spatial regions and varying time periods, arbitrary areal units and annual time periods, and arbitrary areal units and varying time periods. These data quadrant categories are used to locate files to query. A query is interpreted to determine the quadrant containing the requested data file. This file is then accessed.

Chapter 7 consists of considerations relating to the design and implementation of spatial data systems. These considerations are not detailed. They relate to a range or continuum of hardware-software configurations necessary to implement the desired spatial data systems capabilities.

This continuum of hardware-software configurations extends from a general purpose computer and a general purpose programming language to a specifically designed computer configuration and a specifically designed on-line query and programming capability. The advantages and disadvantages of systems configurations along this continuum are discussed with respect to problems and needs for urban data.

## Organization of Part IV

Part IV contains internally unrelated chapters. These are extensions or elaborations of topics developed in the prior chapters. Part IV contains a discussion of linking separately collected data pertaining to the same object or location, a discussion of the organization of data for retrieval,* a discussion of reducing time for searching spatial data, a discussion of techniques for determining whether a point is within an areal unit represented as a polygon, and discussion of retrieving spatial patterns from digitized imagery.

These topics provide superficial treatment of several areas that require considerably more work. Their inclusion focuses attention on their relationship to spatial data systems.

---

*For persons having little or no exposure to data storage techniques, especially list structures, Chapter 9 should be read immediately.

# CHAPTER 2. DATA HANDLING CAPABILITIES FOR SUPPLYING THE DEMANDS FOR URBAN DATA

## Introduction

In this chapter, consideration is given to the supply and demand relationships for urban data. Existing data handling systems are also investigated. These topics are treated together because the demands for spatial data create the pressures resulting in the development of new data handling capabilities. On the other hand, breakthroughs in systems and sensor technology increase data handling and data collection capabilities. These breakthroughs enable an exogenous shift in the supply schedule of data. Thus, the requirements for future data handling systems are not only specified by existing demands, but by new demands created by new capabilities.

The investigation of existing data handling systems is confined to several systems that are used or are capable of being used to process urban data. Of these, Span (Statistical Processing and ANalysis system) from the System Development Corporation [1], Mark III File Management System from Informatics Inc. [15], Colingo (Compile On LINe and GO) from the Mitre Corporation [17], Integrated Data Store from General Electric [6], GIS (Generalized Information System) from IBM [12], and an On-Line Data Management for the Massachusetts General Hospital from Bolt Beraneck and Newman, Inc. [8], are reviewed. Of these only Span is designed in explicit response to the needs of urban data users.

Not discussed here are numerous package programs available to assist in the statistical analysis and display of spatial data. The data handling systems discussed here enable the preparation of input to statistical analysis and graphic display packages. The more general data handling capabilities that are dwelt upon here, enable more extensive use of the statistical and display packages.

Supply of Spatial Data

In quantitative terms, the supply of data is increased by improved technology for data handling and data collection. Improvements in data handling enable more data to be processed in a given period of time or for a given amount of money. The largest impact in terms of the supply of data is in relation to the increased detail or location preciseness of observations on the earth's surface and the increased frequency of observation. More observations per unit area can be anticipated and these observations are going to be made with increasing frequency.

To illustrate this change of supply, view a set of data values as points in a space-time domain. A space-time domain is defined as a four dimensional coordinate system. Two dimensions correspond to the spatial surface of the earth. One dimension or axis corresponds to a value measure for an observed item. The fourth dimension measures time.

Assume the phenomena are houses. For each house, the ground floor area is observed in addition to recording spatial location and time of observation. Each house or entity in this data set can thus be positioned or visualized in the space-time domain. Geographic coordinates relate each house to the earth's surface. The ground floor area determines the position on the value dimension and the value of time determines the position on the time dimension. Increasing the supply of these data is accomplished by making more frequent observations of the houses. Or by observing additional characteristics of the houses additional domains are created for the same set of entities.

As an alternative to considering each house as an entity, one can partition the spatial surface into a set of uniform cells. Then the number of houses in each cell can be observed. In this case, increasing the supply of data can be accomplished by partitioning the area into more and smaller cells as well as increasing the frequency of observation and the properties observed. The appropriate density of data points in the space-time domain is dependent upon the type of phenomena being observed and the use of these data. For example, the frequency of observation of housing units in an area is not as important as the frequency of observation of vehicles in an intersection.

Technology is bringing forth data of more types, observations on a much finer spatial scale, and more frequent observations. The technological advances are of many types, such as aerial surveillance for increased areal scope, resolution and frequency; and new data processing for handling increased numbers of observations.

New sensors and automated interpretation and pattern recognition illustrate the technological advances that promise greater numbers of space-time domains. However, the data output from these exotic data collection techniques are no different than the data as described above. For data are values of properties of entities. Therefore, whether data collection be by conventional means or by new remote sensors the data representation in digital form are the same. Data handling capabilities are therefore common to digital data regardless of the method of collection.

## Demand for Spatial Data

Requirements for data processing systems vary considerably within the area of urban analysis. On one hand are the functional or operating agencies empowered with the management of specified processes. These agencies deal with technical problems, administrative support and fiscal activities. On the other hand, systems for the manipulation of data for the purpose of policy making are becoming more prevalent. The requirements for the two types of systems, i.e., functional systems, and policy guidance systems, are considerably different. The differences accrue from the variance in objectives. Functional systems are oriented to the present, to discrete activities, and to servicing or control. Conversely, policy guidance systems are oriented to the future, to comprehensive activities, and to decision making.*

Both functional systems and policy guidance systems are concerned with and designed on the basis of existing notions or perceptions of data requirements. With impending and existing improvements in hardware technology new levels of data supply capabilities are possible. The data demands must relate to these new data capabilities. Breaking away from constraints upon data requirements imposed by limited supplies of data, new sets of questions can be formulated. These new sets of questions and the new capabilities of data availability open doors to new levels of need. The general availability of new technological capabilities permits better use of greater amounts of data. Thus, technology allows the perception of new classes of problems. Similarly, acceptance of public control over urban environment engenders new sets of questions.

## Evaluation of Existing Data Systems

The following discussion of data handling systems indicates the range of existing systems technology. These systems provide a capability for supplying urban data in usable forms. Some elements of the described systems have not been applied to handling urban data. These features offer new capabilities and enable the creation of new demands for data.

The purpose of this evaluation of existing data systems is to determine the data handling capabilities that exist for the manipulation and retrieval of urban data. These capabilities are searched for desirable features for incorporation into a system of storage and retrieval of spatial-temporal data. As well, the evaluation provides the reader with a concise review of available techniques.

### Span

Span is a system for management of urban data. It is a large-scale data

*See Grundstein [10] or Issacs [11] for a discussion of classifying urban data systems into functional and policy guidance categories.

management, file processing, and statistical analysis system, programmed for the IBM 7090/7094. Span was initiated at the Penn-Jersey Transportation Study in late 1962 and is being completed at System Development Corporation (SDC) with support of the Bureau of Public Roads, U.S. Department of Commerce. Span is generally described in reference [1]. References [2], [3], and [4] are technical memoranda relating to various details of Span.

Briefly, Span is a flexible user-oriented system for the reduction, manipulation, and display of data. Processing functions in Span are "implicitly" programmed. That is, the user communicates with the system through simple directives and parameters stated in English-like sentences. The user thus selects, through the parameters, a particular configuration of pre-programmed procedural options and adapts program operation to his precise requirements. This variation of a package program concept provides considerable flexibility and power. This flexibility and power is due to the large number of options available to the user and because the system is not designed for a restricted set of data types but for data in general terms, i.e., data are values for properties of entities in matrix form.*

The advantage of Span is that it offers an entire system that is capable of handling the bulk of demands by urban analysts. The disadvantage of Span relates to its implementation, but is of little concern as long as IBM 7090/7094 equipment are ubiquitous. However, this will not be the case given the present introduction of third generation computers.

Essentially, Span is technologically bound to the IBM 7090/7094 computer and to the Fortran II monitor system. Conversion of Span to another hardware configuration would require extensive modification, both in terms of rewriting the package programs written in machine language and the Span Systems Supervisor

*Span is restricted to handle only fixed-length records, however.

operating under Fortran II. But the more serious problem in moving to new equipment and monitor systems, is that many Span features are IBM 7090/7094 features that would not be desirable or necessary if implemented on other computer systems.

Span has made some unique contributions to the capability for handling urban data. It is hoped that continued development of this system will result in a more machine-independent data manipulation language rather than a collection of package programs that operate under a supervisor system written for a specific computer system.

Assuming the Span Syntax for Process Control to be a language rather than a means of specifying parameters, translators could be written for a variety of computer systems. And the language could be very similar for different computer configurations.

However, Span is operational and it offers great flexibility and power in accessing and manipulating data. For those persons having access to IBM 7090/7094 hardware a better capability for handling urban data does not exist.

Mark III

Mark III is described in general terms for persons interested in urban analysis in reference [15] and is applied to a typical data base problem in reference [14]. Mark III is of interest to urbanists as an early version of Mark comprises the Metropolitan Data Center computer system [13]. The Mark III system has also been implemented for managing urban data for the city of Alexandria, Virginia. Mark III is designed and implemented for IBM 1400 series hardware requiring a minimum of 12 K core and 4 tape drives [14].

The Mark III File Management system consists of three major subsystems: File Creation and Maintenance Subsystem, Information Retrieval Subsystem and Report Generating Subsystem. Like Span, Mark III is a system of package programs, with the user supplying parameters. Mark III does not have syntax or activated by English-like instructions. Rather, the parameters are entered on forms and converted to control cards.

The File Creation and Maintenance Subsystem creates files on magnetic tapes from punch cards or card images on tape for use by the other subsystems. Mark III has the ability to accomodate a wide variety of magnetic tape files. These files may contain variable length records by means of subrecords. Thus, the first level may contain several data items and a count. The count is automatically developed by the system and indicates the number of sub-records occurring at the second level. Each sub-record may contain a count for an additional level of sub-records. The Information Retrieval Subsystem is a record selection based upon sequential searching or the inspection of every record to see if its data satisfy sets of logical criteria provided by the user. Finally, the Report Generating Subsystem consists of forms to specify the format of reports. Some, though it is not clear how much, arithmetic capability is available in this output package.

Like Span, Mark III is designed and implemented for a specific computer system. However, with the availability of 1401 emulators for the IBM 360 and for other manufacturer's equipment, Mark III has a prolonged life without necessitating reprograming.

Being implemented on the IBM 1401, Mark III does not have the computational power of Span or other systems that are implemented on large-scale computers. But for an economic in-house data processing capability for manipulating urban data, Mark III offers a strong system.

Colingo

A general description of the design philosophy of Colingo is presented in reference [17] and is applied to a data base problem in reference [9], as are extensions Colingo C-10 and Adam. This discussion will essentially be confined to the original Colingo implemented for a 16K IBM 1401 computer with auxiliary disk storage. However, many of the limitations of Colingo are nullified in Colingo C-10 implemented on the IBM 1410 and by Adam implemented on the IBM 7030.

Colingo is designed to refer to data structures through a data dictionary. Therefore, the Colingo Control Language (CCL) does not refer to data fields directly, but through the dictionary with the property identifiers.* In addition, CCL has the power to redefine the dictionary during execution. This enables a greater flexibility in manipulating data.

In CCL, action-verbs and punctuation activates programs that come into use in the sequence determined by the CCL statement. CCL statements have the following capabilities:

1. Data qualification (logical processing using an IF action-verb.)
2. Data computation (mathematical processing, with COMPUTE and SORT action-verbs).
3. Output (PRINT, REPORT, etc. action-verbs).

CCL is both a basic query language and a limited programming language approaching the capability of general purpose programming languages. Although the design objective was to make CCL independent of the computer equipment configuration, core-size limitations of the 1401 has necessitated some compromise. Thus, the Colingo system chains CCL statements and allocates them to disk. Then the statements are sequentially called in by a small in-core executive routine.

The 1401 core, operating under Colingo, has room for only one action program and only one data record (and its describing dictionary) to exist simultaneously. An entire data file is read and processed before Colingo calls the action program corresponding to the next action verb to process these data again. The variable names, logical operators, and arithmetic operations that are imbedded

---

*Dictionaries are created using a Cobol-like data division for format descriptions.

in the CCL statements are the parameters for the action programs.

Colingo is a production system implemented on a small machine for data stored in sequential files. Whereas, extensions of Colingo-Colingo C-10 and Adam are random-access oriented, Colingo's more rigid structure imposes an one-record-at-a-time approach to processing data. Although Colingo's applications are oriented to the defense industry's needs it could serve adequately for manipulating urban data.

Integrated Data Store

Integrated Data Store (IDS) is generally described in reference [6] and is applied to a data base problem in reference [5]. IDS is a programming language to facilitate the organization, storage, maintenance, and retrieval of data using a mass memory storage medium. IDS is designed to be implemented in conjunction with Cobol to provide list processing capabilities that are not specifically available in the Cobol language.

The IDS language enables data description by declaring the existence of master/detail relationships between records as illustrated in Figure 2-1.

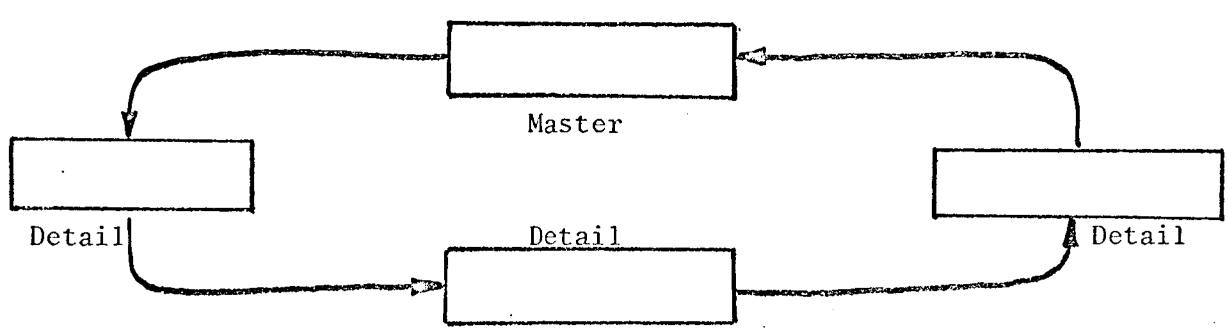



Figure 2-1. Schematic Diagram of Master/Detail

These relationships are implemented through chain link (list processing) techniques which create circular or ring structure. Each record in the chain contains the address of the next.

The IDS language consists of verbs, such as STORE, RETRIEVE, MOVE TO WORKING

STORAGE, and others, that are used within Colbol to actuate list processing functions. The IDS commands automatically create, maintain, and delete the chain structure specified in the data description.

A version of IDS is available for the GE-200 series computers. However, this version is integrated with the general assembly language rather than Cobol, thus making it more difficult to use. IDS/Cobol for the GE-400 and 600 series computers is presently being implemented and offers the full power of the combined language. IDS/Cobol is an extension of Cobol to enable list processing. As such, IDS/Cobol could be implemented generally by computer manufacturers. According to the IDS designers there is no reason why IDS-like capabilities cannot be added to Fortran, Algol and PL/1.

With IDS the data base becomes the center of interest. The computer is just a tool used by the user's program to facilitate the manipulation of the data base. The IDS commands enable both serial and random access to data records.

Other than features incorporated in response to the objectives and clientel for IDS capabilities, there is little new in IDS when compared to previously developed list processing languages.* IDS introduces to business data processing a formal language capability for list processing. Coupled with Cobol, list processing capabilities should prove to be a powerful tool in business data processing.

Previously developed list processors, such as Flip (Fortran LIst Processor), IP-V and Lisp have not gained acceptance in the business data processing community. Whether reinventing the wheel in form of IDS will accomplish acceptance of list processing by the business data processors remains to be seen. The advantages of IDS relate to its emphasis upon the structure of the data base, the actuation

---

*List processing capabilities and languages are introduced in Part IV, Chapter 9 as part of a discussion of structuring data in list form.

of list processing functions by a minimal number of verbs, and the incorporation of the list processing verbs into a general programming language. These advantages are not restricted to business-oriented users and are applicable to users in other areas.

A great deal of know-how has been accumulated, in the development of IDS that is applicable to storage and retrieval of data on urban or spatial data. IDS capability to organize, store, maintain and retrieve complex structures of data, is a much needed capability in handling urban data. How directly IDS will influence the so-called Urban Information System is dependent upon the acceptance of Cobol by the urban systems people. In any event, the implications of list processing is made more evident and applicable by IDS.

GIS

The Generalized Information System (GIS) is an IBM-developed set of programs for System/360.* GIS is designed to perform data file establishment, maintenance, retrieval, and presentation operations common to many application areas. GIS enables the development of a data base in a convenient form for access. It is also an adaptable user-oriented system for manipulating data to fit a wide range of special requirements.

The GIS language consists of a number of variable parameter statements. This allows a user to establish and modify control tables, to enter task specifications, and to control task execution. A wide range of file structures and organizations may be established, maintained, and used. As a generalized data handling capability, GIS is very serviceable for retrieving and presenting data.

The first statement of a GIS program identifies the type of job and specifies the appropriate data file. One of the more important statements in the GIS

---

*GIS is described in [12], which is available from most IBM branch offices.

language is the conditional statement. Conditional statements qualify or select data entities for further processing. Qualifying records are placed on a temporary list for subsequent processing or presentation. Also available are data reduction statements to sort, count, total, and average a list of values, and procedural statements for presentation or output.

It is necessary for the user to define data sets to the system. The description is entered and stored within the system in a permanent library of data description tables. A data file is stored once, allowing tasks to be executed against it without further reference to its storage format.

GIS provides for interrelating data items in different physical data sets. A link allows for locating a higher-level item through a logical key, in the case of indexed sequential data sets, or through a relative address for direct access data sets. For example, linkage may be used to relate person trips from a trip file to the household of a household file that made the trip. GIS also enables development of a chain of subordinate items that are associated with a master item. In the trip-household example, a chain would be appropriate for each household master item. Trips generated by a household are subordinate items on a chain for that household.

GIS promises to be a powerful data handling capability, with a great deal of flexibility to respond to a wide variety of problems and applications. If procedures are incorporated into GIS to handle problems associated with spatial data, GIS will serve these data handling needs very adequately. For example, data reduction statements to determine whether a point is within a polygon, to calculate an area, and to perform simple statistical tests, are essential needs in handling of spatial data.

More so than IDS, GIS offers a generalized data handling capability for the manipulation of complex data structures. Further GIS is not embedded in an existing language (Cobol), which is unfamiliar to urbanists. Upon widespread implementation of GIS, there will be considerable experimentation in its use for urban and spatial data.

On-Line Data Management System

Finally, an on-line data management system for the Massachusetts General Hospital developed by Bolt Beranek and Newman, Inc. is briefly described [8]. This system was developed for research purposes and not for large-scale production. The principal function of the system is to investigate the feasibility of using a computer to handle the real-time medical administrative operations of a hospital. Its relevance to processing urban data is due to on-line retrieval capabilities and because of its specialized design for retrieval from complex data structures.

The system is implemented on a time-shared computer servicing 48 remote teletypes. The hardware configuration of the computer system includes a modified Digital Equipment Corporation PDP-1 computer, a 50-million-character UNIVAC Fastrand drum, two magnetic tape units, and a small 400-thousand-character drum. Only 4,000 words of the 24,000 word computer is available for running a user program. The small drum is used for program storage during multi-programming and the big drum and magnetic tapes are used for bulk storage of data and library programs. The 4K storage for user programs, small buffers, 10 character per-second teletype output, and one-tenth of a second drum access preclude large-scale production.

The one-line data management system is designed for use by the non-programmer. All user programs appear to users in the form of question-answer type of dialogue and provide abundant checking of both semantic and syntactic errors, error messages, facilitates for correcting error easily, and rapid verification of uncoded entries. Use of the system consists of calling a desired program from the library via the teletype terminal. Then the user carries on a dialogue with the computer, with the user-program acting as intermediary. The course of the dialogue, as well as the actual functions performed, are determined by the program, the user, and the current data.

For example, there is a general-purpose retrieval program which accepts any Boolean or arithmetic combination of fields as a selector. A linear search is performed on the file to find records fitting the selector.

The operation of the generalized data input and retrieval system is based on the answers to specific user program questions. There is no need for anything approaching a programming language. Answers, often a simple "yes" or "no," are stored as parameters in the program or in the file itself. The program interprets these answers and performs its function immediately.

The above description is essentially a direct extraction from the description of the system contained in [8]. The purpose of including a discussion of this system is to illustrate the approach to the development of an on-line retrieval system. The on-line orientation engenders different design responses. Another interesting feature of the system is the method of defining a file structure and the details of retrieving from these files.

It is safe to assume that retrieval systems for urban data will move in the direction of real-time capabilities. These capabilities will be realizable via time-shared, on-line systems. The experience as illustrated by Bolt Beranek and Newman's endeavor will prove useful; more useful than generalized on-line programming systems because of the file orientation of the system.

## Status of Urban Data Handling Capabilities

It appears that urbanists are not leaders in adapting to available technology. There are many features of the above described systems that are applicable to processing urban data, but that have not been used. The usable supply and utility of data can be increased by utilization of available systems technology.

There is an immense number of alternatives available. For the urban analyst, the problem is largely one of specifying a set of problems in terms that a systems configuration can be determined. The demands for data, in explicit form, would surely be augmented by creating an efficient and flexible data handling capability. Without such a capability the questions relating to urban problems are not put in explicit demand form.

Adapting capabilities from other areas is not sufficient. Urban data systems are composed largely of data on spatially distributed phenomena. Storing and handling spatial data is unlike the problems of business data processing and scientific computation, and new concepts and new systems approaches are necessary to enable accessing spatial data. A comprehensive discussion of spatial data is presented in the next chapter.

## References

1. Almendinger, Vladmir, "Span: A System for Data Management," Contained in Rogers [16].

2. ______________________, Span Reference Manual: Span System Operation, System Development Corporation, TM-1563/010/01, February 1965, (AD 613284).

3. ______________________, Span Reference Manual: Span Data Transformations and Stratification Capability, System Development Corporation, TM-1563/014/01, March 1965, (AD 613285).

4. ______________________, Span Reference Manual: Data Files Manipulation and Processing, Systems Development Corporation, TM-1563/0121/01, March 1965, (AD 613289).

5. Backman, Charles W., "Integrated Data Store," Contained in Baum and Gorsuch [7].

6. ______________________ and S. B. Williams, "A General Purpose Programming System for Random Access Memories," AFIPS Fall Joint Computer Conference Proceedings 1964, Vol. 26, p. 411.

7. Baum, C. and L. Gorsuch (eds.), Proceedings of the Second Symposium on Computer-Centered Data Base Systems, System Development Corporation, TM-2624/100/00, 1965, (AD 625417).

8. Castleman, Paul A., "On-Line Data Management System for the Massachusetts General Hospital," Contained in Baum and Gorsuch [7].

9. Cataldo, Frank, "Three Colingo-like Approaches to the Data Base Problem," Contained in Baum and Gorsuch [7].

10. Grundstein, Nathan D., "Urban Information Systems and Urban Management Decisions and Control," Proceedings of the Third Annual Conference On Urban Planning Information Systems and Programs, American Society of Planning Officials and Northwestern University, 1965.

11. Issacs, Herbert H., User-Oriented Information Systems for State and Local Government, System Development Corporation, SP1988, 1965, (AD 612997).

12. International Business Machines Corporation, Generalized Information System: Application Description, White Plains, N.Y., 1965.

13. Medin, Anthony C., "Developing a Data System Project from a Concept to Working Operation," Contained in Rogers [16].

14. Postley, John A., "Mark III File Management System," Contained in Baum and Gorsuch [7].

---

*Documents with AD numbers are available from: U.S. Department of Commerce, Office of Technical Services, Washington 25, D.C.

References (Cont'd)

15. Postley, John A., "A General Purpose Urban Management Data System," Proceedings of the Third Annual Conference on Urban Information Systems and Programs, American Society of Planning Officials and Northwestern University, 1965.

16. Rogers, Clark D. (ed.) Urban Information and Policy Decisions, Institute of Local Government, Graduate School of Public and International Affairs, University of Pittsburgh, 1965.

17. Spitzer, J.F., J.G. Robertson, and D.H. Neuse, The Colingo Design Philosophy, The Mitre Corporation, November 1964, (AD 608833).

## CHAPTER 3. NATURE OF SPATIAL-TEMPORAL DATA.

### Introduction

To urbanists the concern is with data on spatially distributed phenomena and relationships between phenomena. The nature of phenomena, places, and relationships in time and space is investigated here. This chapter provides a detailed discussion of spatial-temporal data. It reviews and introduces fundamental concepts concerning the various aspects of data. The purpose of this detailed treatment is to provide a framework for the discussion and development of data systems for the management of spatially distributed urban data.

### Definition of Data

When phenomena from the real world are observed and recorded, data are created. These observations are often quantitatively measured. Quantitative data are collections of values of selected properties of entities. [5] Values are a measurement of a property and are measured on a nominal, ordinal, interval or ratio scale.* An entity is an observation on a phenomenon. Phenomena endure through time, whereas entities are observations upon a phenomenon at a point in time or for some specific time interval. A set of entities or a file is composed of discrete observations. Entities are observations on phenomena having similar characteristics, such as persons employed in a certain industry, persons possessing valid drivers license, or vehicle trips of a specific type. Symbolically $v_{ij}$ is a value for the $j$th property of the $i$th entity. The matrix $V_{mn}$ consists of

---

*See Hodge [3] or Siegel [6] for a discussion of measurement scales for quantitative data. Non-quantitative data must exist in other forms. Welt [8] argues that "information begins with data, which then are given superstructure of language..." Capabilities for processing information in its natural English language form are being developed. Synthex has been developed at SDC to answer English questions posed to a computer-stored file of English text material. [4]

m entities and n properties.

Data are of two types--content data and place data. Content data relate to recorded observations upon phenomena and place data relate to geometric characteristics of segments of space. Content data refers to observations on phenomena that occupy a particular place at a particular time. On the other hand, only properties of size, shape, area, and connectivity of space are considered place data.*

Content data have three elements or dimensions:

1) Phenomena

2) Location

3) Time

Geographic data must possess all three elements. The class or type of phenomena that is to be observed must be specified. Similarly, the location and time of observing phenomena must be specified. For content data, the entity structure is defined by these dimensions.

## Properties of Data

Almendinger [1, p. 16] partitions properties into two types--code and variables. He defines codes as either identifying particular entities or relating entities to one another. Variables are defined to be "primarily observational in origin and represent computation data for an entity." This division of properties does not account for the in-between area, i.e., codes identifying a classification scheme of variables, nor does it account for different types of variables.

For the purpose of this discussion, properties are partitioned into the following types based upon measurement scales:

---

*This terminology is from Thomas [7] and is adopted here to achieve consistency for investigations conducted as part of Contract Nonr 1228(37), Department of Geography, Northwestern University.

1. Location codes identify or locate individual entities or sets of entities. The entities are partitioned into mutually exclusive classes represented by symbols that express equivalence for entities of the same class. This system of classes constitutes a nominal scale. A person's county of residence is a nominal scale location code. It is merely a partitioning of persons by their county of residence. All persons in a county are equivalently located.

   The name method of location identification as used by Barraclaugh [2] is an example of measurement on a nominal scale. The name method of geographic identification is the designation of a name or code for areal units delineated on a map. The map is a necessary adjunct for determining the relationship of areal units from their codes. Census tracts, school districts, and counties are usually referred to by their name or code.

   Geographic identification using a coordinate system is subject to a higher order of manipulation. Location coordinates are values on a ratio measurement scale. Coordinates not only provide unique identification of locations but distances between coordinate values can also be calculated.

2. Classification codes are of three types as distinguished by measurement scale.

   a. Elements of a set may be partitioned into arbitrary Non-ordered Groups identified with codes. These codes or symbols are measurements on a nominal scale. A land use classification is a partitioning of land uses. Occupation and industry breakdowns are non-ordered groupings represented by classification codes. Classification of persons by sex and race also fall in this category, as do vehicle license numbers.

b. Elements of a set may be partitioned into arbitrary Ordered Groups identified with codes. These codes or symbols are on the ordinal scale. For example, classification of occupations, physical conditions, social rank, and ratings may be hierarchical.

c. Data categories refer to a code assigned to class intervals of values measured on a ratio or interval scale. These codes are subject to ranking and the distance between classes is known. Therefore these values are on an interval scale. Examples are the representation of age, income and time categories by codes.

3. Variables are values measured on the interval or ratio scale.

a. Interval variables are properties whose values possess a definable unit of measurement capable of representing the distance between any two values. Interval scales have arbitrary zero points. Examples are scales measuring temperature and time.

b. Ratio variables are properties whose values are additive. Ratio measurements have meaning both for the distance between values and for the ratios of their values from an absolute zero point. Distance and weights are examples.

The above classification of properties is formulated to be consistent with methods of data recording and with methods of measurement. A classification of properties for real-world data is significant in that one can determine the kinds of manipulation of values that are valid. In addition, Siegel [6] discusses the levels of measurement and the statistics appropriate to each level. The power of a value determined by the measurement scale used plays a large role in selecting tools for analysis of data.

## Data Entities

The units for observing phenomena are determined and stated in terms of the dimensions--phenomena, location, and time. Data are observations upon some set of phenomena, constrained by location and time. These data may be organized as discrete observations upon each phenomenon or as aggregates organized by location and/or time.

Content data organized and aggregated by location are commonly called areal data. Typical examples of areal data are census tracts or traffic zones. The actual phenomena of interest may be housing units, households, persons, or trips, but the unit of observation or the entities are areal units. In areal data, location defines the entity structure. The phenomena are secondary to the principal argument of location.

Content data organized and aggregated by time intervals are commonly called time series data. Typically such time intervals as day, week, month, quarter, year and decade are used. Observations, or phenomena, such as employment, population, production, and sales are often aggregated to time intervals for analysis of variation over time. In time series data, the entity is defined by the principal argument of time.

## Place Types

Data on spatial phenomena are related to three types of places or regions. The first type of place is the coordinate region. Secondly, query regions are arbitrary groupings of coordinate regions. The third type of area is a uniform data region. As the name implies this latter class relates to areas of homogeneous function.

Coordinate Regions

A coordinate is one of a set of numbers which locate a point in space. Coordinate regions are an exhaustive partitioning of the study surface into uniform spatial areas whose dimensions equal the fineness or accuracy of the coordinate system. The existence of coordinate regions is necessary for the development of the other two types of areas. Query regions and uniform data regions are defined in terms of coordinate regions.

Query Regions

Query regions are arbitrarily delineated places. These regions are, for the convenience of machine processing, expressed as polygons and are a grouping of coordinate regions. This grouping of coordinate regions or polygon is used to approximate the shape of any conventional areal unit of interest. Administrative, political, analysis, and social areas can be described as polygons. Query regions are often used for collecting and aggregating content data. Census tracts are arbitrary areal units that fall within the class of query regions. Content data related to a query region by testing whether a coordinate pair is within a polygon.* Once the determination of whether data are within a query region the data may be aggregated to form a new entity whose principal argument is location.

Uniform Data Regions

The uniform data region has a very restrictive purpose. The uniform data region is a device to indicate the areal extent of phenomena. Phenomena have locational properties. These properties may be approximated by a single coordinate pair or by defining a region to approximate the extent of a phenomenon.

---

*Methods for determining whether points are in a polygon are discussed in Part IV, Chapter 11.

An example best illustrates the need and function of a uniform data region. A coordinate region of resolution 10 or 100 feet may not entirely include the whole function of a single phenomenon. An industrial plant that covers a number of acres and a number of coordinate regions is usually described as a single entity or observation with a location property of a single coordinate pair. If the region that the industrial plant covers is described as a polygon it is then possible to relate the single entity to any of the included coordinate regions. This is of significance where a query region contains a portion of a uniform data region but not the single coordinate pair that is the conventional location property of the entity. By testing for intersection of the uniform data region and the query region these ambiguous situations are avoided.

Uniform data regions are arbitrary areas that are defined to be homogeneous and relate single phenomena to places. The entity structure defines the criteria for uniform data regions. These may or may not be mutually exclusive. Ownership entities and industrial establishments are examples of entity structures defining what may be overlapping or contained uniform data regions.

## Dimensions of Data

A phenomenon can be considered to occupy a location in n-dimensional space. The dimensions correspond to properties. Observations on a set of phenomena result in a set of points in n-dimensional space. The value of a property determines the distance of the point from the origin on the axis corresponding to that property.

A set or file of data consists of entities having common properties. A set of data has at least one dimension with all points having the same value. This value restricts the location of all points in n-space and each point lies on a hyperplane defined by the common value. For example, all objects extracted from an aerial photo have a common value for time and lie on a hyperplane defined by that time value.

The n-dimensional space concept is an extension of that space-time domain used in discussing the supply of urban data in Chapter 2. The space-time domain is a four-dimensional space, with two dimensions acting as the spatial surface of the earth as represented by a geographic coordinate system, the third dimension a measurement scale for an observed value, and the fourth dimension representing time. Extending this to n-dimensions merely lets the number of observed properties of a thing be variable rather than singular.

To illustrate the dimensions of data the following example is presented. Consider the problem of analyzing the travel behavior of a sample of households. One set of observations is made upon the household itself to assist in the determination of the effect of family composition, occupation, income, etc. upon the family's trip making behavior. In addition, characteristics for each trip may be collected, and analyzed separately or aggregated to the household level for analysis of households. Confining the example to analysis of household trip behavior, the trip characteristics are observed as properties of the household, e.g., total number of trips by type taken on day prior to the interview.

Each observed characteristic of the household is a property or dimension of data. Each household is a data point positioned in n-dimensional space according to the observed values of each property. Values for income category, occupation, and family size determines the point location of each household observation.

## Principal Dimensions

The principal dimensions or elements of content data are further explored. Principal dimensions are those properties that define the class of phenomena to be observed and that limit the extent of the data. Principal dimensions are a select set of all dimensions which are essential for collecting, storing, retrieving, and manipulating data. The principal dimensions or elements of content data are phenomena, location, and time.

By specifying these dimensions, content data are organized to a discrete entity for each phenomenon, areal data, or time-series data. This specification of principal dimensions is done by means of, what is to be called, arguments. With areal data, location is the primary argument. With time series data, time is the primary argument for structuring the phenomena to entities.

Phenomena Dimension

Of all possible things or phenomena that may be observed, subsets must be selected. The more selective one is, in defining a subset of phenomena for observation, the fewer degrees of freedom, or fewer properties are variable. If the definition restricts the type of phenomena that are eligible for inclusion, these defining properties are implicitly common and the number of variable dimensions are effectively reduced. Let this selection of phenomena or definition of data be the phenomena dimension. The phenomena dimension includes properties for which values are held constant. To be eligible for inclusion, a phenomenon must lie on the phenomena dimension.

For example, households are defined by the sample procedure and by the definition of the type of grouping of persons that are eligible as being considered households. Similarly, housing units are defined as a unit for collecting census of population and housing data. These terms are used to define reasonably homogeneous items that possess similar properties to be measured. Because of the basic similarity of housing units, comparative data can be observed. These observations are values for properties common to all or most housing units. Properties such as condition, age, value or rent, and size are relevant to all housing units. On the other hand, non-residential buildings are inadequately observed using these properties. The definition of the housing unit entity excludes these differences. Similarly the properties used to determine household travel behavior may not be suitable for characterizing individuals or commercial travel behavior. The definition of the household insures sufficient homogeneity so that variable values for common properties can be measured.

All data are a subset of a more general class of data. Data are retrieved or a new set of data is created by defining a more restrictive phenomena dimension to source data. By a more restrictive definition, fewer phenomena qualify. More properties become fixed in value rather than variable. Thus, households having income within a specific range, whose head has a specified occupation, and who own two or more cars are a subset of the original sample. These households have a more selective definition. Similarly, dilapidated housing units are a subset of housing units, defined by increasing the properties falling within the phenomena dimension.

## Location Dimension

Most data of interest to urbanists have locational properties. Special consideration is necessary for analysis of data that varies by location. The interest and importance of data on spatially distributed phenomena pose problems to geographers and urbanists that differ from the manipulation of data for business and scientific applications. Thus, the special emphasis upon the locational properties. This emphasis is reflected by defining location dimension as a principal dimension.

The location dimension is a two-dimensional* spatial coordinate system to define positions on the earth's surface. Alternatively, areal units, such as census tracts, school districts, and counties, are implicitly located on a spatial surface by means of codes representing the areal units and a map for relational position. For analysis of data that are spatially distributed, provision

*This concept could be extended to three dimensions.

must be made to let position or location be a variable dimension.

In addition to determining position within an area, the location dimension may be used to define the spatial extent for eligible phenomena. Just as the phenomena dimension defines an eligible set of things, the location dimension is used to define spatial extent. For example, the household travel survey is restricted to some finite region. These data are analyzed to determine the aggregate travel behavior of persons within this region. The location dimension is first used to define the spatial extent of data to some finite region. That is to say, that the households lie within the boundaries of the community or traffic zone being analyzed.

Becoming more prevalent than before is the practice of associating a unique coordinate pair with each household observation. This in addition to defining the eligible region enables the points in n-dimensional space to be represented as being spatially distributed within the eligible region. Without unique location, the location dimension is effectively subsumed in the phenomena dimension. A binary choice exists, phenomena are either within the region or not.

Restated, the location dimension are those properties that 1) define spatial extent in which phenomena must lie to be eligible for observation, and 2) locate phenomena within the region. Using the household travel behavior survey example, the study area boundaries would comprise the region. Coding the survey data to quarter square mile zones would comprise the location within the region. The quarter square mile zones are used as aggregation units. For unique location of discrete households, location precision to about ten foot grids would be desirable. However, for sample data there is little need to carry the location to anything less than quarter square mile zones. At this level, the sample is expanded and the zone must be treated as being homogeneous. Where samples are not used and full-count data are used, the unique location of things within a region are observed and recorded. Rather than treating all observations within a zone as being similarly located, location is a variable which enables positional correspondence of data to the real world.

When phenomena, such as the travel behavior data, are aggregated to quarter-square mile zones or other areal units, location becomes the primary argument for querying these data. Queries to these data must be related to the new entity--the area. For example, for quarter-sections, what are the number of households possessing two cars. Here the area or place is the primary argument.

Time Dimension

An extremely important property of data is the time at which a set of phenomena is observed. Phenomena are observed during a time interval. A file of data usually consists of phenomena observed at essentially the same point in time or observations upon the same type of phenomena at several points in time. An example of the first type of observation is counting vehicles or houses from an aerial photograph. The latter type of observations is exemplified by a set of data on the annual number of motor vehicles registered in some area over a period of years. An example of observations upon phenomena both simultaneously and serially is population figures for each county of a state and for a number of years. The entities are counties with properties being yearly population.

Time is an important part of observing phenomena. Whereas a thing or phenomenon may endure through time an observation upon a thing captures its status at a specific time. Things change as time passes. In order to describe these changes data must be collected at various times. One of the main problems is to identify the same thing in two different time periods.* This usually involves a unique identifier associated with each thing. At each observation point the unique identifier allows relating entities from each set that correspond to the same things. The problem is developed and discussed after introduction of data

---

*See Part IV, Chapter 8 for a discussion of Entity Linkage. Entity linkage is matching the same object from sets of separately collected data.

categories.

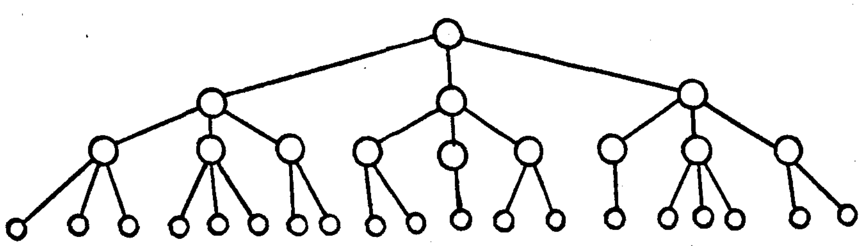
The importance of the time dimension depends on the type of phenomena being observed. Phenomena that rapidly change their properties obviously require more frequent observation than slowly changing phenomena. For example, parking usage studies require frequent occupancy counts to accurately depict usage. On the other hand, annual land age inventories usually adequately reflect changes in land usage. Even the household travel behavior data can be collected over a period of time, say two to four months, and still be considered as being collected simultaneously, for the purpose of transportation planning.

The time dimension is considered a principal dimension because time, as does location and name, is used to define a set of observations. This set is comprised of entities resulting from observing things having similar time properties.

When data on phenomena, such as vehicular accident records or building permit applications are aggregated to intervals of time, time becomes the primary argument for querying these data. Queries to these data must be related to the new entity--the time interval. For example, in a month, how many vehicular accidents have occurred. Here the time interval is the primary argument.

## Argument Hierarchy

To better illustrate the importance of the primary argument, a hierarchal concept of arguments is presented. The argument hierarchy must relate to a hierarchy grouping of data. Data grouped in hierarchal fashion are viewed as illustrated by the tree of Figure 3-1.



| CASE A | CASE B |
|---|---|
| Location | Time |
| Time | Location |
| Age | Age |

Figure 3-1 Hierarchal Grouping of Data

To access data organized in hierarchal groups the arguments must be ordered in the order of the grouping hierarchy. In Case A of Figure 1, the first level below the data universe of say persons, are locational groupings. Below the locational groupings are temporal groups and below there are, say age groups of individuals. To directly access entities, i.e. individual persons, without an exhaustive search, the arguments must be expressed in order of the grouping hierarchy--location, time and age for Case A and time, location and age for Case B.

Type of Content Data

Above, data are classified as content data and place data. Here, content data are subdivided into object slices, events, activities, and transactions. These data are classified and distinguished by their spatial and temporal characteristics. The principal dimensions--phenomena dimension, location dimension, and time dimension--provide the basis for this classification. Primarily, the location dimension and time dimension enable distinctions to be drawn between various data. These data are categorized and differentiated by the principal dimensions.

1. An object slice is an observation at a point in time upon some phenomena situated in space. Examples of object slices are observations on persons, vehicles, street segments, ownerships, households, licenses, and output from image processing (interpreted objects). Object slice data capture the status of a selected set of phenomena at a point in time. Through object slice data, an inventory is taken of some phenomena.

Subclassification of object slice data appears in a following section. Additional classification is necessary to organize the multitude of things that are observable and that fall within the general definition of object slice. Related to this classification is the need for rules for identifying the same object observed in two different time periods.

2. Events are occurrences in time and space. Examples of events are auto accidents, building permit applications, zoning variance approvals, and police and fire calls.

Event data are a collection of observations on occurrences. The events themselves are occurrences in time. The collection of events comprise a file or set. This collection is for a time period, e.g., the vehicle accidents at some location over the period of a year. Event data entities are observations upon discrete occurrences and are stored and analyzed as collections of occurrences within a time period.

3. Activities are cumulative occurrences at a location for a specified duration. Examples of cumulative occurrences are, the annual retail sales of an establishment and hourly vehicle arrivals at a location. Annual retail sales are the cummulation of individual sales or events. Activities are also processes, tasks, or projects. Examples of tasks or processes are characteristics of a construction project, of a printing job, or of any undertaking or action having a duration.

Activity data are characteristics of phenomena that occur over a period of time at a point or in a region. An activity is accomplished or takes place and this process has an associated duration.

4. A transaction is a movement or flow between two points. These flows take place over some time period or duration. Examples of transactions are vehicular flows betweeen two points, person movements between two points, and commodity and financial flows.

Transaction data are observations on movements of thing(s) between two points or regions. The entities in this system may consist of individual things moving from an origin to a destination, or a set of origin-destination combinations with flow characteristics between them.

## Tracing Phenomena Through Time Using Object Slice Data

Inventories or object slice data are very common. Often there is a need to trace phenomena through time using object slice data. Although object slice data are not very suitable for this, there is often no other choice. Essentially, the process involves linkage of the same phenomena as two separately collected data sets, i.e. time slices. The means for performing this linkage depends on the type of phenomena. Phenomena that are static can be linked by a locational argument, whereas a dynamic phenomena require unique identifiers to enable linkage of the same individual or object from two different data sets.*

Observations on the different types of phenomena may take various forms. An image or photograph represents object slice data most accurately with respect to time. Yet objects must be identified through an interpretation process. Interpretation is the selection of patterns possessing specified characteristics. These patterns are then said to represent a particular kind of objects.

### Static Phenomena

When comparing static phenomena, such as streets, at two points in time, and identical patterns are found at the same location, the patterns are assumed to represent the same object. This assumption is reasonably valid as long as the time interval between images is short enough to prohibit the demolition of one object and the construction of another. Naturally, the maximum allowable time interval varies according to the kind of fixed object being observed. The allowable time interval for a tree would be longer than that for a building. Replacing

*Entity linkage is discussed in greater detail in Part IV, Chapter 8.

a tree with one of similar size, would take a number of years for growth. Most buildings can be demolished and replaced within a year. In summary, fixed objects observed at the same location in both time t and time t + 1 are assumed to be identical, or the same object. It is further assumed that the object existed throughout the time interval t to t + 1, that is, if the time interval is less than the maximum allowable interval for that kind of fixed object.

Dynamic Phenomena

Using object slice data, dynamic phenomena can only be traced through time if they are uniquely identifiable. For example, a vehicle observed at location $\ell$, time t, cannot be assumed to be at location $\ell$ at time t + 1, where one unit of time is sufficient to move the vehicle. Given vehicle v, situated at $\ell$ at time t. At time t + 1 vehicle v can only be found by a unique identifier. If it is possible for a vehicle to move a distance r in one time unit, the area $\pi r^2$ centered on $\ell$ must be scanned looking for vehicle v.

In general, dynamic phenomena must be uniquely recognizable in order to trace them through time using object slice data. Object slice data captures the characteristics of phenomena at a point in time. To trace dynamic phenomena it must be identified in two time periods. If the object has changed position the assumption is that it has moved directly from point $\ell$ to point $\ell'$.

Although object slice data are most easily visualized as imagery they can appear in other forms as well. Inventories and many record keeping systems are object slice data. These capture the status at a point in time. An inventory records the stock on hand at a given time. Similarly, the Census of Population is an inventory of living objects.* Record-keeping systems, such as vehicle license files, are updated periodically, whether it be daily or monthly. Annually, the

*The lack of unique identification prohibits tracing individual family units. Only aggregative measures of migration are possible.

file is completely updated using new vehicle license applications. Such a file can be considered an image or inventory of the current status of the phenomena the records attempt to capture.

Imagery is little used to trace human movements except for isolated studies of small spatial areas. Tracing person movements is of two types--micro-movements or macro-movements. Micro-movements are a persons movements to and from work, shopping, school, etc. The aggregate daily flow of persons are data of interest to transportation planners. These data are not usually collected in object slice form. Rather these data are in the form of flow or transaction data. On the other hand, macro-movements are movements such as change of residence, work place, locality, or region.

## Data Organization

Efficient and flexible organization of spatial-temporal data is the key to development of meaningful information system capabilities for applying spatial data to urban analysis. It is essential that data be organized in a form that enables transformations to be made in response to a variety of needs or demands. Both a flexible form of data and a capability to manipulate data must exist.

Data organization needs are largely underemphasized, yet it is important that data be organized to increase their utility. Data should be organized in forms capable of responding to a wide variety of questions and capable of transformation to a variety of analysis forms.

A tutorial discussion of data organization is presented in Part IV, Chapter 9. This discussion provides a background in file searching, data redundancy, list structures, and associative memories. It also suggests means of organizing spatial data.

Essentially, the need is to respond to demands. These demands relate to phenomena, location, and time. Questions concerning phenomena specify locations and time. Variations to questions are formed by letting phenomena, location and/or time be variable. A wide variety of questions may be posed by letting one, two, or three of the dimensions be variable and the remainder fixed or specified. For example, by letting phenomena be unknown, one can ask--What phenomena are at location $\ell$ during time interval $\Delta t$? By letting phenomena and location be unknowns, one can ask--What phenomena move from location $\ell$ to all other locations during time interval $\Delta t$?

The problem is to translate questions such as these to queries that relate to available or obtainable data. Chapter 4 is concerned with this problem.

## References


1. Almendinger, Vladimer B., SPAN Reference Manual: Data Files Manipulation and Processing, System Development Corporation, TM-1563/021/01, (March 1965), (AD 613289).

2. Barraclough, Robert E., "Geographic Aspects of Information Retrieval," Urban Information and Policy Decisions, Graduate School of Public and International Affairs, University of Pittsburgh, 1965.

3. Hodge, Gerald, "Use and Mis-use of Measurement Scales in City Planning," AIP Journal, Vol. 29, No. 3, (May 1963), pp. 112-121.

4. Issacs, Herbert H., et. al, Natural Languages Computer Proceedings of Los Angeles Police Department Crime Information, Progress Report No. 1, System Development Corporation, TM-1793/000/00, (April 1964).

5. Language Structure Group of the CODASYL Development Committee, "An Information Algebra: Phase I Report," Comm. of ACM, Vol. 5, No. 4, (April 1962), p. 190.

6. Siegel, Sidney, "Nonparametric Statistics," American Statistician, Vol. 11, No. 3, (July 1957), pp. 13-19.

7. Thomas, Edwin, Geographic Information Systems: Final Report, Nonr 1228 (37), Department of Geography, Northwestern University, (Forthcoming).

8. Welt, Issac D., "User Requirements and Public Relations," Conf. Tech. Inf. Center Admin., Vol. 1, Arthur W. Elias (ed.), Drexel Inst. Tech., Phila., Pa., (June 1964), Spartan Books, Wash. D.C., 1964, p. 171.

# CHAPTER 4. QUESTIONS AND QUERIES.

## Introduction

Providing answers to questions directed to an information system requires the questions be expressed in a form relatable to data. Expression of a question in data relatable form is defined as a query. Whereas a question is a user's expression of need, a query is a transformation of this question to selection criteria. First, an appropriate set of data is selected by the query, then appropriate entities are selected from this set. A query is merely an imperative statement that requests data satisfying some specified criteria.

In utilizing an information system, the basic problem is to translate questions to queries. The user expresses the question. The information system translates this question to a query or set of queries which select data to provide an adequate response to the original question. It may be that the available data provides only a partially satisfactory answer. In these cases, the partial answer is input to a subjective human process to answer the initial question.

The statement that the information system translates the question to a query does not necessarily imply the process is automated. Viewing an information system as a man-machine system, translation may be a human process, i.e., an essential part of an information system is a question translating technician. The logical process of retrieval is illustrated in Figure 4-1. This process merely differentiates between the user and system. It does not specify that elements of the systems are automated.

As illustrated in Chapter 3, data are of diverse types. Retrieval of these data necessitates a flexible query system that enables accessing data by specification of principal dimensions--phenomena, location, and time, and selection

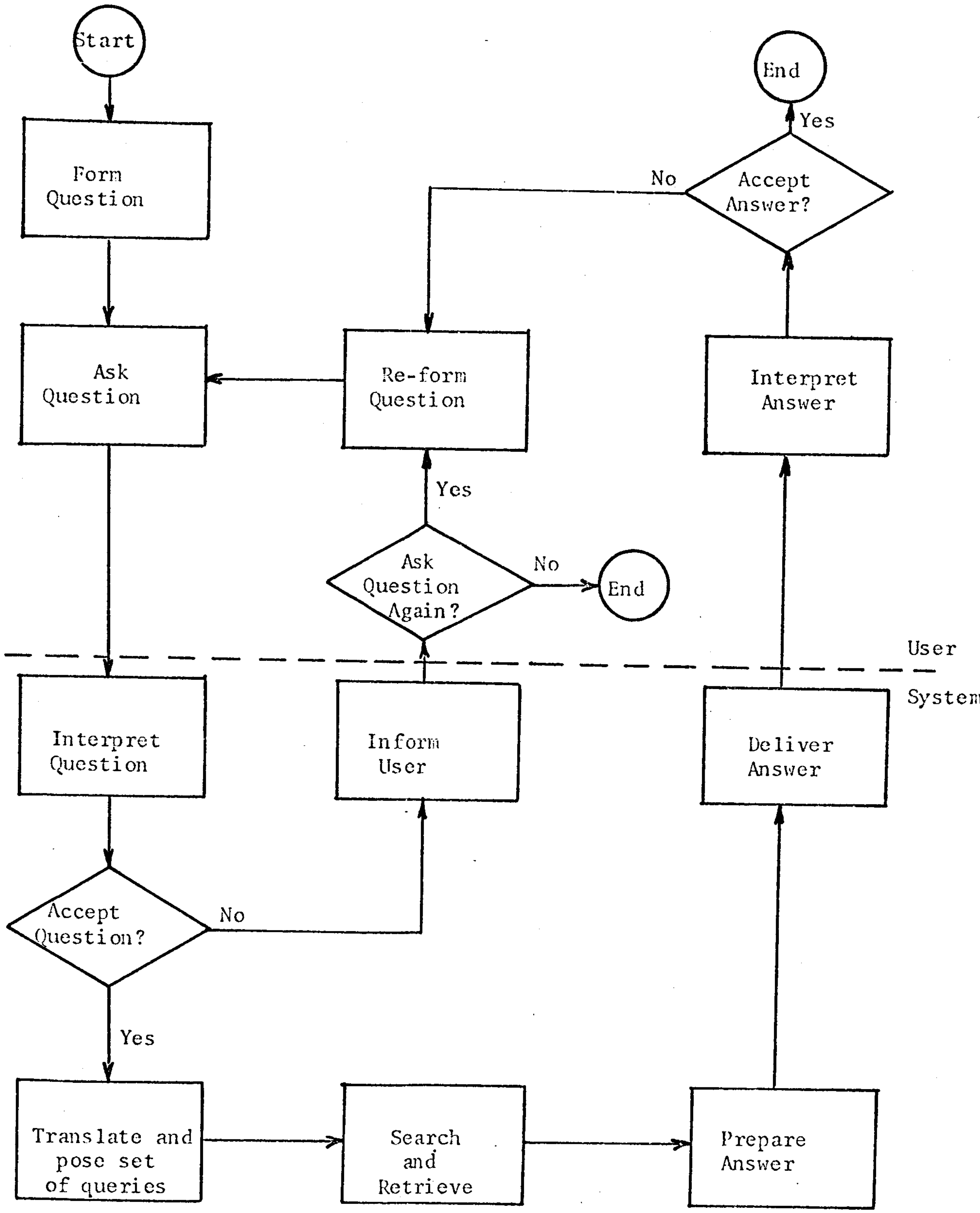



Figure 4-1. Logical Process of Retrieval

(Adapted from Blunt [ 1 ])

of entities by specifying values for properties. Thus, a general query capability is desired* rather than concentrating on translation of natural language questions for a restrictive class of data.** A generalized data handling or query capability is realizable in the short run, whereas a question answering system is a longer range task.

Questions

Statements are transformed into questions by means of the question mark, intonation, and the rearrangement of subject. In addition, special question words--who, why, where--are used to specify the nature of the desired answer. The question itself consists of two elements. One delineates a set of alternatives which include the answer and the other requests a particular subset of these alternatives [4].

The set of alternative elements may be disjunctive, i.e., offers a choice between yes and no or true and false. The request is to choose the correct alternative. For example, "Is the intersection clear of vehicles?" can be answered yes or no. "Who is in the intersection?" requests, from all possible vehicles, the identification of those within the intersection. In addition, questions with built-in assumptions such as "Is the intersection still congested?" or "Is the intersection still empty?" are often posed. Questions of this type assume a previous condition or assume that a term, such as "congested" is unambiguous.

Questions of concern for analysis of urban systems relate to characteristics of objects, events, activities, and flows that exist in urban regions. In

---

*A query language possessing these characteristics is presented in Part III, Chapter 5.

**Question-answering systems that have been developed only deal with small subsets of English and what data of limited variability. See Farrell [2] and Green [3] for examples of retrieving information via natural language questions. See Simmons [4] for a detailed review of experimental English language question-answering computer systems.

addition, there are questions as to phenomena characteristics in a place or time interval. In general, three classes of questions may be identified. One class relates to phenomena, where the phenomena dimension is the principal argument; the second relates to phenomena, where the location dimension is the principal argument; and the third relates to phenomena, where the time dimension is the principal argument.

Since this study is not concerned with automatic question answering system, little formalization of questions themselves is necessary. Rather, the concern is with formulating queries.

## Queries

A query was defined above as a question in a form relatable to data. Relating a question to data, transforms the question to an imperative command. The imperative command specifies the selection of a set of entities or file(s), and selects entities possessing specified characteristics from the set.

At the first level, a query is directed to a set of data. By the subject matter of the question, a query is formulated to a given data set. With a question relating to condition of housing, a query is directed to places whose characteristics are housing conditions or to a file of characteristics of houses. If a question is posed for which data are not available, a query cannot be formulated.

At the second level of query formulation, is the specification of subsets or subfiles of data. Urban data are often partitioned into subsets according to locational and temporal characteristics. For example, like data are often organized into files according to city, state or other areal units, and are organized into files by year, month, day, or other temporal groupings. Specification of these locational and temporal properties that are used to organize data into files enables the selection of a more appropriate set of alternatives.

This second level of query formulation is of great importance when data are organized by phenomena dimension and subfiles exist for various values of the location and time dimensions. Means of formulating these types of queries are presented as part of Quest in Part III, Chapter 5. A more detailed discussion of grouping data into spatial-temporal sets is also presented in Part III, Chapter 6.

Once the appropriate file or subfile is selected, entities may be selected. Selection of entities may be according to locational, temporal or other criteria. However, the means of selection are essentially the same. A query, as an imperative statement, calls for selecting entities from appropriate file(s). A query specifies characteristics that an entity must possess to be selected. Characteristics are specified by defining allowable value ranges for properties and relationships between properties. For example, selection of households possessing characteristics of owning two or more cars, income of greater than $10,000 per year, and family size of equal to or less than three. The properties of interest are number of cars owned, total family income, and family size. The value limits for these properties are specified and the relationship between these properties, in this case all must be satisfied. In computer programming such a selection criterion is easily set up as a Boolean expression within an IF statement.

Finally, the entities selected by the query must be identified. This may be done by direct reporting, or some manipulation or processing of these selected entities may be desired. Discussion of manipulation, modification, and output alternatives are left to the discussion of Quest in Part III, Chapter 5.

More complex queries are necessary when questions are posed as to the relationship between places or phenomena. Such questions may be translated to queries directed to data on relationships such as transaction data. However, data does not exist for many relationships. Relationships must be inferred from comparing places, objects, or by linking data relating to the same object at two points in time or from two separately collected data files.* For example, the relationship between the characteristics of school children and family composition characteristics might be derived by directing queries to a file of school age children and to a file of characteristics on families, matched on last name and street address.

Given the state-of-the-art of automatic question-answering systems, as reviewed by Simmons [4], flexible question-answering systems do not appear imminent. Well defined and often asked questions may be automated, but the task of accessing data in unanticipated ways is beyond present automated question-answering capabilities. Thus, the effort of this study is to develop a query answering capability. At this stage of technology, translation of questions to queries is a human process. Considerably more study of translation of questions to queries is necessary before automating the process. It is not done here. Rather, the next chapter describes a language to query spatial-temporal data.

*See Part IV, Chapter 8 for a discussion of entity linkage.

## References


1. Blunt, Charles R., An Information Retrieval System Model, HRB-Singer, Inc., State College, Penna, October, 1965, (AD 623590).

2. Farrell, Jules, A Natural Language Information Retrieval System, System Development Corporation, TM-2392, May, 1965, (AD 615763).

3. Green, B.F., Jr., et. al., "Baseball: An Automated Question Answerer," Computers and Thought, E. A. Feigenbaum and J. Feldman (eds.), McGraw-Hill, 1963, p. 207.

4. Simmons, R. F., "Answering English Questions by Computer: A Survey," Comm. of ACM, Vol. 8, No. 1, January, 1965, p. 53.

Unclassified
Security Classification

## DOCUMENT CONTROL DATA - R&D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

| | |
|---|---|
| 1. ORIGINATING ACTIVITY *(Corporate author)*<br>Department of Geography<br>Northwestern University<br>Evanston, Illinois. | 2a. REPORT SECURITY CLASSIFICATION<br>Unclassified<br>2b. GROUP |

3. REPORT TITLE

SPATIAL DATA SYSTEMS: ORGANIZATION OF SPATIAL DATA

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(Last name, first name, initial)*

Dueker, Kenneth J.

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| December, 1966 | | 29 |

| | |
|---|---|
| 8a. CONTRACT OR GRANT NO.<br>Nonr 1228 (37)<br>b. PROJECT NO.<br><br>c. Task No. 389-143<br><br>d. | 9a. ORIGINATOR'S REPORT NUMBER(S)<br>Technical Report No. 4<br><br>9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* |

10. AVAILABILITY/LIMITATION NOTICES

The distribution of this document is unlimited.

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| | Geography Branch<br>Office of Naval Research |

13. ABSTRACT

Spatial data systems are concerned with the organization, handling, and retrieval of data whose spatial position is of concern. Spatial data are of particular concern in urban and transportation planning. In these fields considerable attention is given to spatial and temporal variations of data.

The three volume report presents a discussion of concepts and techniques that are essential in moving towards flexible and responsive urban information systems. The following areas are emphasized:

1. Explication of terms associated with spatial data.
2. Discussion of means of organizing spatial data for flexible and efficient retrieval.
3. Investigation of data handling capabilities for organizing and manipulating spatial data.
4. Presentation of these topics in a tutorial form, conceivably to serve as a text where none presently exist.

The greater speeds and storage capacities of newer computers requires new concepts of data organization and new means to create and access these more complex data structures. Of particular concern in urban and transportation planning are needs to link separately collected data that relate to the same phenomena or spatial locations, and a need for user-oriented data handling capabilities. These needs are explored and recommendations are made.

Unclassified
Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| Geography | | | | | | |
| Spatial Data Systems | | | | | | |
| Urban Data | | | | | | |
| Language | | | | | | |

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.